

The Role of Bodily Perception in Emotion: In Defense of an Impure Somatic Theory

LUCA BARLASSINA* AND ALBERT NEWEN

Ruhr-Universität Bochum

Abstract

In this paper, we develop an impure somatic theory of emotion, according to which emotions are constituted by the integration of bodily perceptions with representations of external objects, events, or states of affairs. We put forward our theory by contrasting it with Prinz's (2004) pure somatic theory, according to which emotions are entirely constituted by bodily perceptions. After illustrating Prinz's theory and discussing the evidence in its favor, we show that it is beset by serious problems—i.e., it gets the neural correlates of emotion wrong, it isn't able to distinguish emotions from bodily perceptions that aren't emotions, it cannot account for emotions being directed towards particular objects, and it mischaracterizes emotion phenomenology. We argue that our theory accounts for the empirical evidence considered by Prinz and solves the problems faced by his theory. In particular, we maintain that our theory gives a unified and principled account of the relation between emotions and bodily perceptions, the intentionality of emotions, and emotion phenomenology.

1. Introduction

In this paper, we propose that emotions are constituted by the integration of bodily perceptions with representations of external objects, events, or states

* Corresponding author. Address: Ruhr-Universität Bochum, Institut für Philosophie II, GA3/39, Universitätsstr. 150, D-44780 Bochum, Germany. Tel.: +49-234-32-24721. E-mail address: luca.barlassina@gmail.com (L. Barlassina).

of affairs.^{1,2} In the simplest case, emotions are multimodal perceptual states in which bodily perceptions are integrated with perceptual representations of external objects. In the most sophisticated case, emotions are cognitive states resulting from the integration of bodily perceptions with propositional attitudes.

We shall develop our account by contrasting it with Prinz's (2004) theory, according to which emotions are constituted by bodily perceptions only. The fact that we criticize Prinz and put forward our theory as an alternative to his, however, hides two major points of agreement between Prinz and us: first, both our theories are *naturalistic theories* of emotion; second, both our theories qualify as *somatic theories* of emotion. In this introductory section, we outline what distinguishes a naturalistic theory of emotion from an a priori one (§1.1); we then explain what a somatic theory of emotion in general amounts to and point out the crucial difference between Prinz's *pure* somatic theory and our *impure* somatic theory of emotion (§1.2).³

1.1. *Naturalistic vs. A Priori Theory of Emotion*

A nice illustration of the distinction between naturalistic and a priori theories of emotion is due to Griffiths (1997). He correctly notices that mainstream philosophy of emotion has almost entirely relied on conceptual analysis: philosophers have usually built their theories of emotion through making explicit the descriptions that competent speakers implicitly associate to emotion terms (Kenny, 1963; Solomon, 1984). For example, in order to develop a theory of fear, philosophers have typically considered how the competent speaker applies the word 'fear' to possible cases and have attempted to articulate the descriptions that speakers have in mind (Davis, 1988). There is nothing surprising with this a priori methodology. This is how analytic philosophy usually proceeds. However, Griffiths contends that, at least in the case of emotion theory, this methodology is seriously flawed.

We maintain that Griffiths is right. Indeed, the term 'emotion' and the names of individual emotions are natural kind terms which purport to refer to kinds of psychological states. However, there are overwhelming reasons to maintain that the reference of a natural kind term is not fixed by the descriptions that competent speakers associate to those terms. Rather, a natural kind term directly refers to the natural kind that was picked out when the term was introduced via an ostension of one of the samples of the natural kind in question (Putnam, 1975; Kripke, 1980). Accordingly, the meaning of, say, 'fear' is not given by the descriptions that speakers have

¹ The notion of bodily perception is explained in section 2.1.

² Hereafter, we use the less cumbersome expression 'representation of external objects.'

³ The distinction between pure and impure somatic theories is taken from Prinz (2004).

in mind. Rather, ‘fear’ refers to the psychological state in which people are in certain paradigm cases. But the nature of these psychological states cannot be discovered by merely inspecting our concepts. Therefore, we subscribe to Griffiths’s conclusion that “if philosophers want to know about emotion, rather than about what is currently believed of emotion, analysis must proceed hand in hand with the relevant empirical sciences” (Griffiths, 1997, 6).

We label a theory of emotion that is not derived from conceptual analysis but rather built on empirical evidence a ‘naturalistic theory of emotion.’ Both our theory and Prinz’s qualify as naturalistic theories of emotion in this sense. Two things follow from this. First, in this paper we shall discuss a large number of experimental data from psychology and neuroscience, and we’ll argue that our theory explains these data better than Prinz’s. Importantly, however, we will not restrict empirical evidence to the findings of cognitive science. In fact, we maintain that an adequate theory of emotion also has to account for *facts* concerning the intentionality and the phenomenology of emotion.⁴ Also in this case, we will argue that our theory does a better job than Prinz’s. The second consequence of adopting a naturalistic stance towards emotion theorizing is that the resulting theory might be somewhat in contrast with our pre-theoretical intuitions concerning emotions. This is the case, indeed. Like Prinz, we endorse a somatic theory of emotion, and such a theory is to some extent at odds with the folk-psychological understanding of emotions. Let’s consider this issue in greater detail.

1.2. *Pure and Impure Somatic Theories*

What is a somatic theory of emotion? How does it differ from a non-somatic one? To a first approximation, a somatic theory of emotion is a theory that maintains that the body plays a crucial role in emotion. This initial characterization, however, faces two difficulties in distinguishing a somatic theory from a non-somatic one. The first difficulty concerns the notion of body. According to an ordinary interpretation of the notion, the brain is part of the body. However, if ‘body’ refers to the brain too, then more or less all theories of emotion turn out to be somatic theories,⁵ since it is a platitude that the brain is crucially involved in mental states and processes, emotions

⁴ We of course do not naïvely presuppose that all phenomenological experiences have to be taken at face value. Their significance is evaluated in the context of all our knowledge of the phenomenon.

⁵ More precisely, if ‘body’ is taken to refer to the brain too, then practically *all naturalistic theories of emotion* turn out to be somatic theories. On the other hand, an a priori theory might be non-committal about the relation between emotion and the brain, or might even deny any connection between the two.

included.⁶ To avoid the risk of trivialization, we adopt a narrower notion of body, according to which ‘body’ refers to all body parts with the exception of the brain.⁷ This, of course, doesn’t mean that a somatic theory of emotion maintains that the body, but not the brain, plays a crucial role in emotion. Rather, the idea is the following: the body, *in addition* to the brain, plays a crucial role in emotion.⁸

However, this refined characterization will not suffice to single out somatic theories either—and here is the second difficulty—since the vast majority of theories of emotion acknowledge that emotions involve the body in this narrower sense. More precisely, most theorists endorse the idea that emotional episodes *cause* bodily changes to occur. For example, disgust is typically defined as the emotion that is caused by perceiving rotten food, excrement, unpleasant odors, etc., and *causes* visceral reactions, nausea, a characteristic facial expression, etc. (e.g., Rozin et al., 2008). What is special with a somatic theory of emotion is that it reverses the cause-effect relationship between emotions and bodily changes: it is not emotions that cause bodily changes; rather, it is bodily changes that play a causal contribution in emotion generation. This, of course, brings us back to William James’s idea that we do not cry, strike, or tremble because we are sorry, angry, or fearful, but “we feel sorry because we cry, angry because we strike, afraid because we tremble” (James, 1890, 449-50). It is in this respect that somatic theories of emotion depart from the folk psychological intuitions about the nature of emotion. Indeed, common sense seems to suggest that emotions cause bodily changes, and not the other way around.

To make a long story short, the distinctive tenet of a somatic theory of emotion is the following:

- (i) Bodily changes (narrowly construed) play a causal contribution in emotion generation.

Aside from (i), disagreement looms large among somatic theorists. A first important distinction is that between pure and impure somatic theories of emotion. According to a *pure somatic theory*, (i) bodily changes play a causal contribution in emotion generation, and (ii) emotions are *entirely constituted* by the perception of such bodily changes. On the other hand, an *impure somatic theory* accepts (i), but denies (ii), and maintains instead that

⁶ This point was made, in a different context, by Goldman and de Vignemont (2009).

⁷ Analogously, we stipulate that the expressions ‘bodily processes,’ ‘bodily changes’ and their cognates don’t refer to processes and changes in the brain.

⁸ Importantly, proponents of somatic theories of emotion disagree about which specific set of bodily changes (narrowly interpreted) causally contribute to emotion generation. We are going to touch upon this issue in due course.

(ii*) the perception of bodily changes is *just one* of the constituents of emotions.

As is well known, a pure somatic theory of emotion has been independently proposed by William James (1884, 1890) and by Carl Lange (1885), and has been considered as outright implausible throughout the 20th century. To begin with, Cannon (1927) argued that emotions cannot just be perceptions of bodily changes, because bodily changes alone do not suffice to distinguish either one emotion from the other or emotions from non-emotions. Secondly, philosophers have claimed that a pure somatic theory of emotion cannot account for the intentionality of emotions. Consider my fear of that dog. It is customary to say that such an emotional episode has two types of intentional content (Kenny, 1963): a particular object (that dog) and a formal object (the property of being dangerous). However, if emotions are nothing but perceptions of bodily changes, how can they exhibit such intentional features? Finally, even if one were willing to accept that basic emotions such as fear or disgust are exhausted by bodily perceptions, it seems that the James-Lange theory is totally ill-suited to account for higher cognitive emotions, such as guilt and shame.⁹ In fact, these latter emotions appear to include propositional attitudes among their constituents.

Here comes Prinz. His book *Gut Reactions* (2004) is an impressive attempt to rescue a pure somatic theory of emotion. By considering a huge amount of empirical evidence and elaborating on Dretske's (1981; 1986) psychosemantics, Prinz engages in a sustained defense of the claim that emotions are nothing but perceptions of bodily changes. Even though we consider Prinz's theory to be the most ingenious and sophisticated pure somatic theory of emotion available, we are going to argue that it is nonetheless beset by insurmountable problems. However, this doesn't mean that we shall suggest abandoning a somatic approach to emotions altogether. In fact, we'll give several arguments against Prinz's pure somatic theory of emotion, and we shall show how such arguments can be met by an *impure somatic theory*. More precisely, we shall argue that emotions are constituted

⁹ The distinction between basic and higher cognitive emotions is surrounded with controversy. To begin with, it has been argued that the very distinction is unwarranted (Ortony and Turner, 1990). Moreover, even those that accept the distinction disagree as to which emotions count as basic and which one do not (Ekman, 1999; Izard, 2007; Panksepp, 2007). Addressing these problems would require a paper of its own. For the sake of simplicity, we shall adopt Ekman's (1972) classical list of the 6 basic emotions, i.e., surprise, fear, happiness, sadness, anger and disgust, while we will treat guilt, shame, and jealousy as paradigmatic types of higher cognitive emotions. In section 4.2.3, we will outline a provisional, twofold account of the distinction between basic and higher cognitive emotions. However, elaborating a full-fledged account of the distinction between basic and higher cognitive emotions does not fall within the purposes of this paper.

by the integration of bodily perceptions with representations of external objects (events, or states of affairs).

A second point of contention among somatic theorists concerns the specific set of bodily changes (narrowly construed) that causally contributes to emotion generation. Lange proposed that emotions are caused by changes in the vasomotor system only. James held a more liberal view, according to which the causes of emotions encompass expressive motor behaviors (e.g., facial expressions), goal-directed motor behaviors, and the activity of the Autonomic Nervous System.¹⁰ Prinz adopts an even more inclusive view, since he maintains that the bodily changes that causally contribute to emotions include “states of the respiratory system, circulatory system, digestive system, musculoskeletal system, and endocrine system” (Prinz, 2004, 5). At least in this respect, we side with Prinz. Thus, the relation of our theory to Prinz’s can be summarized as follows. First, both theories are somatic theories of emotion, since they both claim that (i) bodily changes play a causal contribution in emotion generation. Second, both theories endorse an inclusive view concerning which bodily changes play a causal contribution in emotion generation—basically, all the bodily changes and processes that take place outside the Central Nervous System. However, while Prinz’s theory is a *pure* somatic one, in that it accepts that (ii) emotions are *entirely constituted* by the perception of bodily changes, our theory is an *impure* somatic one, since it rejects (ii) and maintains that (ii*) the perception of bodily changes is *just one* of the constituents of emotions. The aim of the paper is to establish that our impure somatic theory has several advantages over Prinz’s pure somatic one.

The paper is structured as follows. We start by presenting Prinz’s pure somatic theory (§2). We then raise several problems for it (§3). Finally, we propose our impure somatic theory of emotion, and we show how it can both solve the problems for Prinz’s theory and make sense of the empirical evidence that he considered (§4).

2. Prinz’s Theory of Emotion

This section is devoted to presenting Prinz’s theory of emotion. We shall begin by illustrating the idea that emotions are entirely constituted by perceptions of bodily changes (§2.1); we then dwell on Prinz’s account of the intentionality of emotions (§2.2); finally, we consider how Prinz deals with higher cognitive emotions (§2.3).

¹⁰ This interpretation of James’s theory has been proposed in Ellsworth (1994). Reisenzein et al. (1995) disagree, and argue that James held that emotions are primarily caused by the activity of the Autonomic Nervous System only. Adjudicating this interpretative issue is beyond the scope of our paper.

2.1. A Pure Somatic Theory: Emotions as Perceptions of Bodily Changes

Prinz holds a pure somatic theory according to which emotions are perceptions of bodily changes. What does this exactly mean? According to Prinz, a mental state is a perceptual state if and only if it is a state in a dedicated input system (or sense modality). For example, visual states are perceptual states because they inhabit a dedicated input system, namely vision. Prinz maintains that emotions too are states in a dedicated input system: “they are states within [the] system that registers changes in our bodies”¹¹ (Prinz, 2004, 58)—recall that, according to Prinz, the expression ‘changes in our bodies’ ranges over the whole variety of physiological changes, from changes in visceral organs to changes in skeletal muscles, from changes in hormone levels to changes in temperature.

Given that Prinz thinks that the perceptual system responsible for registering all bodily changes is the *somatosensory system*, he formulates his point as follows: “in saying that emotions are perceptions of bodily changes, I mean ... to say that they are states within our somatosensory system” (Prinz, 2004, 58). Prinz’s formulation of his thesis, however, is not satisfactory. Indeed, it has been argued in contemporary neuroscience that the perceptual system responsible for registering one’s bodily condition is better characterized as the *interoceptive system* (Craig, 2003), which is mediated by two different pathways: (i) the insular pathway (lamina I spinothalamocortical system), in which afferent fibers that innervate all tissues of body project through the spinal cord and the brainstem to the insular cortex (Craig, 2002); and (ii) the somatosensory pathway, in which skin afferents project to the somatosensory cortex (Khalsa et al., 2009). Many questions concerning the interoceptive system remain unanswered. In particular, the exact contribution of each pathway to interoception has to be determined, and it is still controversial whether interoceptive states—i.e., the registrations of one’s physiological condition—are realized only at the cortical level (Craig, 2009) or also at the sub-cortical level, e.g., in the brainstem (Damasio et al., forthcoming). For the sake of simplicity, in this paper we will conceive of *interoceptive states* as the cortical registrations of the different bodily states that are realized either in the insular cortex (via the insular pathway) or in the somatosensory cortex (via the somatosensory pathway). On this basis, we can rephrase Prinz’s pure somatic theory as follows: emotions are bodily perceptions in the sense of being interoceptive states.¹²

Even though the idea that emotions are perceptions of bodily changes was originally developed by James, there is a crucial point on which Prinz

¹¹ In section 2.2, we shall clarify Prinz’s distinction between *registering* and *representing*.

¹² Hereafter, we will use ‘bodily perception,’ ‘perception of bodily changes,’ and ‘interoceptive state’ interchangeably.

departs from him: while James held that emotions are conscious perceptions of bodily changes, Prinz claims that the bodily perceptions constituting emotions can occur either consciously or unconsciously.¹³ However, Prinz accepts that when emotions are conscious their phenomenology is exhausted by feelings of bodily changes.¹⁴ He also follows James in considering this alleged phenomenological evidence as supporting a pure somatic theory of emotion. Indeed, he argues that the best explanation of the fact that emotional experiences are exhausted by bodily feelings is that emotions are entirely constituted by perceptions of bodily changes.

Prinz resorts to other two lines of evidence to establish his pure somatic theory of emotion. First, he presents empirical findings showing that voluntary and involuntary changes in people's facial expressions influence their emotional reactions towards eliciting emotional stimuli. For example, Strack et al. (1988) asked participants to evaluate the funniness of a series of cartoons while holding a pen either in their mouth or in their hand. Some subjects were asked to hold the pen with their lips, thus inhibiting muscle activity associated with smiling (*lip condition*); other subjects were instructed to hold the pen between their teeth, thus facilitating smiling (*teeth condition*); finally, some subjects held the pen in their non-dominant hand, which did not affect their facial muscles (*hand condition*). Subjects in the teeth condition rated the cartoons as funnier than subjects in the lip condition, while the ratings of subjects in the hand condition fell between the two extremes. According to Prinz, this shows that the subject's "emotional response was being elevated by their unintended facial expressions" (Prinz, 2004, 36), as should be the case if emotions are perceptions of bodily changes.

A second type of evidence comes from neuroscience. On the one hand, Prinz considers neuroimaging studies that revealed activation, during emotional episodes, in the somatosensory and insular cortex (Damasio et al., 2000), i.e., in the neural correlates of interoception. On the other hand, he points at neuropsychological evidence showing that damages to the anterior insula cause deficits in emotional experience (Hennenlotter et al., 2004). Prinz maintains that these neuroscientific data are best explained by the hypothesis that emotions are perceptions of bodily changes.

Prinz is aware that if one endorses a pure somatic theory of emotion, then one has to face Cannon's (1927) objection: emotions cannot be identified with perceptions of bodily changes, because bodily changes do not suffice to distinguish either one type of emotion from the other or emotions

¹³ Therefore, Prinz's theory counts as a pure somatic theory, but not as a feeling theory of emotion.

¹⁴ Prinz has an interesting, if extremely speculative, story concerning how and when the brain gives rise to emotional feelings. For our purposes, however, it is not relevant to dwell on it.

from non-emotions. For example, Cannon claimed that fear, anger, chilliness, hypoglycemia, asphyxia, and fever all involve the same bodily changes. Prinz's reply to this charge is twofold. First, he specifies that he is not committed to the thesis that an emotion type is always brought about by the same physiological responses. Rather, he maintains that each emotion type is associated with a set of bodily changes (a body state pattern) and it occurs when a sufficient number of the members of the set are registered by the interoceptive system. Second, he refers to empirical evidence concerning the six basic emotions (happiness, sadness, fear, anger, disgust, surprise) showing that, even though no single physiological dimension is unique to any of them, each type of basic emotion is indeed associated with a specific body state pattern (Levenson et al., 1990).

Finally, a pure somatic theorist has to provide a principled account of the distinction between the interoceptive states that are emotions and those that are not. Indeed, even if one grants that each emotion type is associated with a specific body state pattern, it remains to be explained why only certain bodily perceptions are emotions, while other bodily perceptions are not. For example, if both fear and fatigue are bodily perceptions, why is only the former an emotion? Prinz's proposal to single out emotions among interoceptive states is as follows: emotions are the only interoceptive states that represent core-relational themes. In order to understand this point, we have to consider Prinz's account of the intentionality of emotions.

2.2. The Intentionality of Emotions

If I am sad about the death of a child, it's implausible to say that my sadness represents the bodily state I am in. Rather, my sadness seems to have the following two kinds of intentional content: a particular object (the child's death) and a formal object (loss). However, if emotions are bodily perceptions, how can they have such intentional features? Let's consider Prinz's answer to this problem.

To begin with, Prinz maintains that it is an error to conceive emotions as representing particular objects:

Arguments for the claim that emotions have intentionality often appeal to the fact that emotions can be directed at some particular event. ... This is a flawed form of argument. While there is a sense in which emotions are directed at particular events, that does not mean that they represent those events The events are represented by mental states that combine with emotions. When I am sad about the death of a child, I have one representation of the child's death and I have sadness attached to that representation. The sadness doesn't represent the death. ... Sadness represents the loss of something valued. (Prinz, 2004, 62)

In other words, Prinz accepts that emotions can be directed towards particular objects, but denies that this means that emotions represent such particular objects. Rather, particular objects are represented by mental states that combine with emotions. More interestingly, Prinz also denies that emotions represent one's bodily condition. In fact, he maintains that emotions only represent formal objects, and follows Lazarus (1991) in giving a naturalistic account of what formal objects are: they are *core relational themes*, i.e., organism-environment relations that pertain to the organism's well-being. Prinz puts it this way:

Sadness represents loss. ... Loss is not a bodily condition. Nor is it something purely external. Loss is a relational property. It is the elimination of something valued by an organism. ... This point applies equally to other emotions. Fear ... represents the property of being dangerous, ... [which is] a relational property. Something can be dangerous only to some creature or other, and whether or not something is dangerous depends on the creature in question. (Prinz, 2004, 63)

Accordingly, this is the problem that Prinz has to solve: how can emotions be perceptions of bodily changes and, at the same time, not represent one's bodily condition, but rather core relational themes? Since Prinz's solution largely relies upon Dretske's (1981; 1986) psychosemantics, let us say a few words about Dretske's theory.

To make a long story short, Dretske argues that a satisfactory theory of mental representations should keep the notions of information and representation distinct. In order for a state *S* to represent *X*, it is not enough for *S* to carry information about *X*; it should also be the case that *S* has the *function* of carrying such information. Since Dretske claims that (a) a state *S* carries information about *X* if and only if *X* reliably causes *S* to occur, and (b) having the function of carrying information about something amounts to having been set in place (i.e., learned or set up by evolution) for that purpose, we can summarize his theory of mental representations as follows:

A mental state *S* represents *X* if and only if: (*Condition 1*) *S* is reliably caused by *X*; (*Condition 2*) *S* has been set in place (by learning or evolution) to carry information about *X*.

On this basis, Prinz develops his account of the intentionality of emotions. Needless to say, Prinz holds that emotions are reliably caused by bodily changes, and hence carry information about bodily changes. However, he argues that emotions do not represent bodily changes, since it is unlikely that evolution set them in place to carry such information: "evolution chooses things that confer a survival advantage ... [and] it is not clear why it is advantageous to know when my blood vessels are constricting" (Prinz, 2004, 59).

Thus, another candidate for what emotions represent is needed. Prinz individuates an alternative candidate by correctly noticing that the bodily changes that reliably cause emotions are in turn reliably caused by core relational themes. For example, fear is reliably caused by a pattern of bodily reactions that are in turn reliably caused by facing a danger. Thus, emotions carry information about core relational themes too. Moreover, such information clearly confers a survival advantage; so, it is a plausible hypothesis that emotions have been set in place by evolution to carry such information. On this basis, Prinz concludes that emotions represent core relational themes.

Prinz further details his proposal by making reference to his dual theory of content (Prinz, 2000). Simplifying a little, he maintains that (at least some) mental representations have two types of content: a real and a nominal content. Consider the concept DOG. This concept represents dogs and dogs only. To be a dog, a creature must have a particular genome. Therefore, DOG represents a particular genome. This is the real content of DOG. Needless to say, speakers have no idea how to describe such a genome. Nonetheless, their concept DOG represents the dog genome. Therefore, DOG doesn't represent the dog genome via descriptions. Rather, DOG has the real content it has in virtue of being reliably caused by the dog genome and having been set in place to carry information about the dog genome. This, however, doesn't mean that the concept DOG doesn't describe any features of dogs. In fact, in addition to its real content, DOG also has a nominal content, which is constituted by the perceptual properties through which we detect dogs, e.g., having fur, barking, wagging their tails, etc. According to Prinz, the same thing applies, *mutatis mutandis*, to emotions. Consider fear. Fear represents its real content, i.e., danger, in virtue of being reliably caused by it and having been set in place to carry information about it. At the same time, fear also has a nominal content. Which one? Since Prinz maintains that fear is a bodily perception, it's natural for him to conclude that the nominal content of fear is a set of bodily changes. In a nutshell, Prinz's story about the dual content of emotions is as follows: an emotion *represents* a core relational theme (*real content*) through a perceptual *registration* of bodily changes (*nominal content*).

Importantly, Prinz uses his account of the intentionality of emotions to distinguish between those bodily perceptions that are emotions and those that are not. According to Prinz, this distinction is a semantic one:

Emotions ... do not represent bodily states. They use bodily states to represent organism-environment relations. Fear, for example, ... has bodily states as nominal contents, not real contents. Now consider fatigue. Fatigue represents a bodily state; it has the function of detecting insufficient rest. Fatigue has a bodily state as its real contents and as its nominal contents. (Prinz, 2004, 190)

In other words, Prinz maintains that what sets aside emotions from non-emotional interoceptive states is that only the former represent core relational themes. More precisely, all interoceptive states *register* bodily changes, and thus all interoceptive states have bodily changes as *nominal content*. However, while emotions *represent* core relational themes, and thus have core relational themes as *real content*, non-emotional interoceptive states represent what they register, and hence they have bodily changes both as real and as nominal content.¹⁵

2.3. Emotions, Cognition, and Calibration Files

One might accept that Prinz's pure somatic theory works for basic emotions such as fear and disgust, but argue that it does not generalize. In fact, there are emotions such as guilt and shame—i.e., higher cognitive emotions—that seem not to reduce to perceptions of bodily changes, but have propositional attitudes among their constituents. For example, it can be maintained that, in order for an emotional episode to be an instance of guilt, it must be at least partly constituted by the belief, say, that one has committed a harmful transgression. Let's consider Prinz's reply to this objection.

Prinz is happy to acknowledge that both basic and higher cognitive emotions can have propositional attitudes among their *causes*. Importantly, however, he denies that propositional attitudes enter the constitution of any emotion. In order to get his point, let's begin with basic emotions. Consider fear. According to Prinz, the bodily changes characteristic of fear can be triggered by different types of elicitors. Sometimes, what causes these bodily changes is just a non-conceptual perceptual state—e.g., the perception of

¹⁵ Prinz not only maintains that emotions represent core relational themes, but also that “emotions are *perceptions* of core relational themes” (Prinz, 2004, 232, emphasis added). An anonymous referee raised the question of how this is compatible with characterizing Prinz's theory as a pure somatic one. In order to answer this question, it is important to note that Prinz explicitly distinguishes two senses of ‘perception’: “Perceptions can be defined in two ways. The first ... has to do with their implementation. Perceptions must occur in perceptual systems. The second has to do with their informational properties. Perceptions pick up information in a distinctive way” (Prinz, 2004, 224). Now, when Prinz writes that “emotions are *perceptions of bodily changes*” (Prinz, 2004, 58, emphasis added), he is using ‘perception’ in the first sense. In other words, he is claiming that “emotions are states within systems that are dedicated to detecting bodily changes” (Prinz, 2004, 224). In this sense, Prinz's theory counts as a pure somatic one, since we have characterized a pure somatic theory as a theory according to which emotions are states within the interoceptive system. The very same states, however, can also be characterized in terms of their informational properties. Since Prinz maintains that emotions pick up information about core relational themes in the way which is distinctive of perceptual states, he concludes that “having an emotion is literally *perceiving our relationship to the world*” (Prinz, 2004, 240, emphasis added).

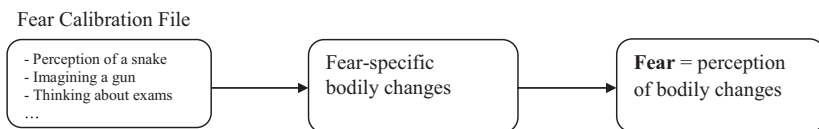


Figure 1. According to a pure somatic theory of emotion, fear is the perception of the bodily changes that are caused by the representations contained in the fear calibration file. More generally, any emotion E is the perception of the bodily changes that are caused by the representations contained in the E calibration file.

a snake-like object (LeDoux, 1996).¹⁶ However, such bodily changes can also be caused by perceptual states with conceptual content, by propositional attitudes, etc. For example, I might undergo an increase in heart rate, intense sweating, etc., by thinking about tomorrow’s exam. In a nutshell, evolution and learning set in place the *fear calibration file*, a data structure containing representations whose activation triggers the pattern of bodily responses characteristic of fear. The representations in the fear calibration file can be either propositional or non-propositional. In both cases, however, they cause fear to occur, but they do not constitute fear. In fact, fear is entirely constituted by the perceptions of the bodily changes triggered by the fear calibration file (Figure 1).

What about higher cognitive emotions such as guilt, jealousy, shame, *Schadenfreude*, indignation, and pride? Prinz maintains that they are *cognitive recalibrations* of basic emotions.¹⁷ For example, he argues that guilt is the result of the cognitive recalibration of sadness. Let us explain. By setting in place the *sadness calibration file*, evolution and learning “calibrated” sadness to be triggered by a certain class of stimuli. Sadness, however, can be “recalibrated,” i.e., its specific bodily changes can be reconfigured to be set off by other kinds of things. According to Prinz, guilt emerges when the bodily changes specific for sadness come under control of the *guilt calibration file*, a data structure that contains beliefs about harmful transgressions. The fact that this new calibration file contains beliefs about harmful transgressions accounts for the intuition that guilt essentially involves beliefs about harmful transgressions. However, such beliefs about harmful

¹⁶ Prinz’s argument to the effect that perceptions of snake-like objects are non-conceptual is as follows. LeDoux (1996) showed that when we see a snake-like object—i.e., a coiled object—the retinal image sends a signal into the thalamus (which cannot perform object recognition), and the thalamus sends then a signal directly to the amygdala, which in turn orchestrates the bodily and behavioral responses that are characteristic of fear. If the thalamus and the amygdala can initiate a full-fledged fear response without engaging the cortex, this shows that such a fear episode occurs independently of cognition.

¹⁷ To be precise, Prinz describes two ways in which higher cognitive emotions are generated from basic emotion. One is cognitive recalibration. The other is blending. Given our aims, however, we skip the reference to the latter.

transgression cause but do not constitute guilt. As in the case of basic emotions, guilt is constituted by the perception of bodily changes only.

It is important to notice that, since the guilt calibration file hijacked the sadness calibration file, the two files are associated with the same pattern of bodily changes. Therefore, sadness and guilt are constituted by the perception of the same bodily changes. Thus, how to distinguish between them? Here is Prinz's answer: when these bodily changes are caused by the sadness calibration file, we have sadness; when they are caused by the guilt calibration file, we have guilt. In other words, even though calibration files are not constitutive parts of emotions, they play a crucial role in determining the identity of emotions.

In a nutshell, Prinz maintains that the difference between basic emotions and higher cognitive emotions is that the latter are cognitive recalibrations of the former: a higher cognitive emotion emerges when a new calibration file hijacks a basic emotion. However, for both basic emotions and higher cognitive emotions, calibration files are just causes, not constituents. Therefore, for Prinz basic emotions and higher cognitive emotions are states of the same type, namely, bodily perceptions under the causal control of calibration files. This allows Prinz to reject a widely held thesis in emotion theory, i.e., the so-called *disunity thesis*, according to which emotion is not a natural kind, but splinters into two or more kinds (Griffiths, 1997). Prinz considers two accounts of natural kinds. According to the first one, all members of a natural kind share a common underlying essence (Kripke, 1980; Putnam, 1975). The second account is due to Boyd (1991; 1999) and is more liberal, since it only requires that members of a natural kind tend to share a cluster of properties in virtue of a homeostatic mechanism, i.e., a causal mechanism that brings about the co-occurrence of such properties. On the basis of his account of the relationship between basic emotions and higher cognitive emotions, Prinz concludes that emotion is a natural kind "in both the Boydian sense and in the sense ... [of having] a unifying essence" (Prinz, 2004, 81). Emotion is a natural kind in the Boydian sense because basic emotions and higher cognitive emotions share a cluster of properties in virtue of being underwritten by the same causal mechanism, which we have depicted in Figure 1 above. Emotion is a natural kind in the essentialist sense because both basic emotions and cognitive emotions share a common essence: they all are bodily perceptions. More precisely, they all are bodily perceptions that represent core relational themes.

3. What Prinz's Theory Cannot Explain

The account of emotion elaborated by Prinz can generate three types of reactions. First, one might be convinced by his pure somatic theory of emotion, and accept that emotions are entirely constituted by bodily

perceptions. Second, one might think that Prinz is on the wrong track entirely, and argue that bodily perceptions are not among the constituents of emotions at all. Third, one might adopt an intermediate stance and embrace an impure somatic theory of emotion, according to which: (i) bodily changes play a causal contribution in emotion generation, and (ii*) the perceptions of such bodily changes are *one* of the constituents of emotions. In the remainder of the paper, we are going to argue in favor of this latter stance. This section has two aims. First, showing that there is robust evidence for the claim that bodily changes causally contribute to emotions. Second, establishing that there are good reasons for doubting that emotions are identical with bodily perceptions.

3.1. Somatic vs. Non-Somatic Theories

As was said earlier on, Prinz considers the following evidence from cognitive science:

- i. Each type of basic emotion has its own pattern of associated bodily changes (Levenson et al., 1990).
- ii. Neuroimaging studies reveal activation in the neural correlates of interoception, i.e., somatosensory and insular cortex, during emotional episodes (Damasio et al., 2000).
- iii. Changes in facial expressions influence one's emotional reactions towards eliciting emotional stimuli (Strack et al., 1988; Zajonc et al., 1989).
- iv. Neuropsychological evidence shows that damages to the insular cortex cause deficits in emotional experience (Hennenlotter et al. 2004).

Prinz argues that (i)-(iv) are best explained by a pure somatic theory of emotion. Is he right? Let's see.

Importantly, (i) and (ii) *per se* do not support a somatic theory of emotion over a non-somatic one, since a non-somatic theorist is not forced to maintain that the body plays no role whatsoever in emotion. In fact, a non-somatic theorist can be perfectly happy to acknowledge that emotions involve bodily changes and bodily perceptions, insofar as such bodily changes are effects and not causes of emotions. And the idea that bodily changes are effects of emotions is entirely compatible with (i) and (ii). Indeed, one might say that each type of basic emotion is associated with a specific pattern of bodily changes because each type of basic emotion

causes a specific pattern of bodily changes. Moreover, since emotions cause bodily changes to occur, this is why interoceptive brain centers get activated during emotional episodes—that is, emotions cause bodily changes, which are subsequently perceived; therefore, bodily perceptions are effects and not constitutive elements of emotion. In a nutshell, what we get from (i) and (ii) is just evidence for a correlation between emotions and bodily changes (and bodily perceptions), which can be interpreted either in the sense that bodily changes cause emotions (as somatic theorists claim) or in the sense that emotions cause bodily changes (as non-somatic theorists claim).

In order to adjudicate between somatic and non-somatic theories, we have to consider (iii) and (iv). This time, the non-somatic theorist is in trouble. Indeed, (iii) shows that changes in one's bodily condition do influence one's emotional state, while (iv) establishes a causal link between bodily changes and emotion by revealing a dependence of emotions on perceptions of bodily changes. Accordingly, there are good reasons to endorse the main tenet of somatic theories of emotion: bodily changes causally contribute to emotions, and not the other way around.

Importantly, however, to say that (i)-(iv) favor a somatic theory of emotion over a non-somatic one is not to say that (i)-(iv) are best explained by a pure somatic theory with respect to an impure somatic one. In fact, in section 4, we will argue that the evidence considered by Prinz can also be explained by our impure somatic theory. In the remainder of this section, we are going to illustrate the difficulties faced by a pure somatic theory. We shall give four arguments against Prinz's theory: it gets the neural correlates of emotions wrong (§3.2); it is not able to distinguish between emotions and bodily perceptions that are not emotions (§3.3); it cannot account for emotions being directed towards particular objects (§3.4); and it cannot account for emotion phenomenology (§3.5).

3.2. *The Neural Correlates of Emotions*

If Prinz is right in saying that emotions are bodily perceptions, it should not just be the case that the brain areas responsible for interoception are activated during emotional episodes. Rather, it should also be the case that emotions and interoception have exactly the same neural correlates—how could it be otherwise if emotions *are* interoceptive states? In order to assess whether a pure somatic theory gets the neural correlates of emotions right, we are going to focus on a specific type of emotion, namely disgust. We choose disgust because it is the embodied emotion *par excellence*. To begin with, it is evolutionarily rooted in a food-rejection system preventing the oral incorporation of bad food (Rozin & Fallon, 1987). In addition, it involves bodily changes such as nausea, retching, and dysregulated gastric responses (Stern et al., 1989; Jokerst et al., 1997; Gianaros et al., 2001; Rozin et al., 2008). Finally,

and more importantly, disgust crucially depends on the interoception of such bodily changes (Harrison et al., 2010). Thus, if we can establish that Prinz's pure somatic theory cannot account for disgust, this *a fortiori* establishes that it cannot be a viable general theory of emotion.

Let's consider the neural correlates of disgust. True enough, disgust is "located" in the insular cortex. However, while the insular correlate of interoception is the posterior insula (Craig, 2002), disgust is "located" in the anterior insula (Krolak-Salmon et al., 2003; Small et al., 2003; Wicker et al., 2003). More precisely, the relationship between the insular region devoted to interoception, i.e., the posterior insula, and the insular region responsible for disgust, i.e., the anterior insula, appears to be as follows. Lamina I afferent fibers project to the posterior insula, where a representation of the physiological condition of all bodily tissues is obtained (Damasio et al., 2000; Craig, 2002). These bodily representations are then mapped onto the anterior insula, where they are integrated with information coming from many brain areas, among which amygdala, cingulate, ventral striatum, and prefrontal and multi-modal sensory regions (Mufson and Mesulam, 1982; Craig, 2002; Harrison et al., 2010). It's only at this point that disgust emerges.

Therefore, even if neuroscientific evidence on disgust favors a somatic theory of emotion over a non-somatic one, for it supports that disgust causally depends on interoception, it is not compatible with the claim that disgust is identical to an interoceptive state. In fact, the neural correlate of disgust is not the same as the neural correlate of interoception. In other words, our first criticism to Prinz's theory can be summarized as follows: it makes a wrong prediction about the neural correlates of emotions.

3.3. *Distinguishing Emotions from the Interoceptive States That Are Not Emotions*

If one maintains that emotions are interoceptive states, then one has to explain how to distinguish emotions from the interoceptive states that are not emotions. Consider fear and fatigue. According to a pure somatic theory, they are both bodily perceptions. However, it's clear that fear is an emotion, while fatigue is not. What is it that accounts for this distinction? As we know, this is Prinz's answer: emotions are distinct from non-emotional interoceptive states in virtue of their intentionality. Indeed, while non-emotional interoceptive states have bodily changes both as nominal and as real content, emotions have bodily changes as nominal content but core relational themes as real content—where a core relational theme is an organism-environment relation that pertains to the organism's well-being. In other words, while non-emotional interoceptive states register and represent bodily changes, emotions register bodily changes but represent core relational themes. Let's call this the *distinctiveness thesis* (DT):

(DT): Emotions are interoceptive states, and what distinguishes emotions from non-emotional interoceptive states is that emotions are those interoceptive states that represent core relational themes (i.e., organism-environment relations that pertain to the organism's well-being).

Recall that Prinz resorts to Dretske's theory of mental representations to account for what emotions represent. Indeed, he endorses the following *psychosemantics thesis* (PT):

(PT) A mental state represents a core relational theme (i.e., an organism-environment relation that pertains to the organism's well-being) if it is reliably caused by it and has been set in place to carry information about it.

However, the conjunction of (DT) and (PT) generates a serious problem for Prinz. Consider, e.g., the feeling of ear pressure. This is an interoceptive state that is reliably caused by an organism-environment relation that pertains to the organism's well-being, namely finding oneself in an environment with unhealthy degrees of barometric pressure and oxygen levels. Moreover, since it is extremely useful for an organism to know whether it is facing such a hostile environment, it is reasonable to hypothesize that evolution set up the feeling of ear pressure to carry such information. If so, we can conclude, on the basis of (PT), that the feeling of ear pressure is an interoceptive state that represents a core relational theme. What's wrong with that? Now, there's nothing problematic *per se* in the idea that the feeling of ear pressure represents a core relational theme. However, if one also endorses (DT), then one has to conclude that ear pressure is an emotion. But the idea that ear pressure is an emotion is absurd.¹⁸ Therefore, Prinz has to give up either (PT) or (DT). In section 4, we will argue the best option available to a somatic theorist is to stick with (PT) and get rid of (DT). But this, *pace* Prinz, will amount to abandoning the idea that emotions are interoceptive states.

3.4. *Emotions Are Directed towards Particular Objects*

Emotional episodes are often (maybe always) *directed towards* particular objects (events, or states of affairs). If John is afraid of that barking dog, then John's fear is directed towards that barking dog. If Mary is worried about tomorrow's exam, then Mary's worry is directed towards tomorrow's exam. If I am happy that you are here, my happiness is directed towards the fact that you are here. The fact that emotions are directed towards

¹⁸ Notice that we are not denying that the feeling of ear pressure can give rise, or be accompanied by, emotions. For example, one might undergo fear when feeling a higher degree of ear pressure. However, the feeling of ear pressure in itself cannot be plausibly considered an emotion.

particular objects is usually explained in terms of emotions representing them: an emotional episode is directed towards a particular object in the sense that it *represents* (among other things) that particular object. For example, John's fear is directed towards that barking dog because it represents (among other things) that barking dog.

Prinz disagrees. That is, he accepts that emotional episodes are usually directed towards particular objects. However, he denies that we have to account for directedness in terms of representation. According to Prinz, in fact, emotions do not represent particular objects at all. Thus, the following two questions arise: first, why does Prinz maintain that emotions do not represent particular objects? Second, can one account for the directedness of emotions if one gives up the idea that they represent particular objects? Let's examine these questions.

First, even though Prinz is not explicit about it, his motivation for denying that emotions represent particular objects is as follows. Prinz's account of the intentionality of emotions is based on Dretske's psychosemantics: a state *S* represents *X* if and only if *X* reliably causes *S* to occur, and *S* has been set in place to carry information about *X*. However, such a proposal cannot be used by a pure somatic theorist to account for how emotions represent *particular objects*. Consider my fear of that barking dog. According to Prinz, such an emotional episode is identical to a perception of a pattern of bodily changes. However, the kind of bodily perception that constitutes fear is reliably caused not only by seeing that barking dog, but also by seeing another barking dog, or by seeing a lion approaching, and it has been set in place to carry information about any of these items. For this reason, Prinz gives up the idea that emotions represent particular objects, and claims that they just represent formal objects, i.e., core relational themes.

Let's now consider our second question: can one account for the directedness of emotions without resorting to the idea that emotions represent particular objects? Prinz's answer is affirmative. More precisely, he puts forward the following account: emotions, i.e., bodily perceptions, are directed towards particular objects because they are *linked* to other mental states that represent those particular objects. Here is an example by Prinz: "If I am sad about the death of a child, I have one mental representation that corresponds to the child's death and another, my sadness, that corresponds to there having been a loss. Together, we can think of these as constituting a complex representation that means that the child's death has been a loss to me" (Prinz 2004, 62-63).

Is this a viable account of the directedness of emotions? To answer this question, it is important to clarify which notion of *linking* is at stake here. Prinz oscillates among three different characterizations of the relation between emotions and mental states that represent particular objects. We are going to argue that they all fail in accounting for the directedness of emotions. Therefore, no matter what characterization Prinz favors, his theory runs into trouble.

3.4.1. First Proposal: Co-occurrence

Sometimes, it seems that Prinz conceives of the link between emotions and representations of particular objects as mere *co-occurrence*: “the neural representation of an emotional bodily state fires at the same time as the neural realization of the representation of its particular object” (Prinz 2004, 181). The following case, however, shows that co-occurrence does not suffice for directedness.

Suppose that I see a barking dog running in my direction and that, as a result, I am afraid of that dog. Prinz would analyze such an episode as follows. I saw a dog; the perceptual representation of the dog triggered a certain set of bodily changes in me; my interoceptive system perceived such bodily changes. Why was that a case of fear of the dog? Because the bodily perception that constitutes fear co-occurred with the representation of the dog. The problem is that such bodily perception co-occurred with many other neural representations. Indeed, my brain also represented the best course of action to choose (e.g., fight, flee, keep calm, look for help, etc.), the best way to perform the chosen action, the appropriate motor program to implement my plan, and so forth. In order to do so, my brain resorted to memory and “looked for” relevant information about my previous encounters with dogs. In addition, neural representations of objects in the surroundings were co-activated as well—e.g., the representation of a tree on which I can jump, the representations of objects that I can use to defend myself from the dog, etc.

In a nutshell, if one conceives of the link between emotions and representations of particular objects in terms of co-occurrence of neural representations, it turns out that an emotional episode is directed towards a myriad of things. But this is not the case. Even though my fear of the dog co-occurred with many representations, it was directed only towards the dog.

3.4.2. Second Proposal: Causation

In other passages, Prinz cashes out the idea of a link between emotions and representations of particular objects in terms of *causation*: “Saying that my sadness is about the death [of a child] does not mean that my sadness represents the death [of a child]; rather it means that the death [of a child] is what has caused me to become sad” (Prinz, 2004, 62). This second proposal fits well with the previous case of the fear of the dog. I perceived the dog, this perceptual representation triggered a set of bodily changes, and I interocepted these bodily changes. Accordingly, my interoceptive state, i.e., my emotion, was directed towards the dog because it was the representation of the dog that caused my bodily changes.

It’s not clear, however, how such a hypothesis can account for cases in which there is a divorce between what triggers one’s bodily changes and

what one's emotional episode is directed towards. Consider the following two stories:

Pete the Runt

Louise and her husband Eric enter a bar. Eric goes to a table, while Louise heads for the counter. The bartender is momentarily absent, and Louise is waiting for his return with another customer—call him 'Alex.' Seated at the table, Eric is reading a newspaper, when he distinctly hears a male voice coming from the counter, uttering obscenities to Louise. This triggers a set of bodily changes in Eric: he feels his pulse quickening, his face reddening, etc. While he's having these bodily feelings, Eric turns to the counter and sees Alex near to Louise. Thus, he gets up from the table, runs towards Alex and pounces on him. Eric is really angry at him: he takes Alex by the throat and shouts: "Get out of here!" Alex, however, is innocent. The obscenities were uttered by a short, nasty ruffian hidden behind the counter, the infamous Pete the Runt.

The Disgusting Can of Coke

Mary participated in a psychological experiment in which she had to look at pictures projected on a screen. During the task, Mary's bodily changes were monitored by different devices. Immediately after the task, Mary had to report what she felt towards the projected objects. In the first trial, Mary was presented with an emotion-eliciting stimulus, namely the picture of a roaring lion. In the second trial, Mary was presented with an emotion-neutral stimulus, i.e., the picture of a mug. Unsurprisingly, observing the picture of the lion triggered fear-related bodily changes in Mary, while observing the mug didn't generate any significant change in her bodily state. In addition, while Mary reported to feel afraid of the lion, she said to have no emotional attitude towards the mug. Finally, Mary was presented with another emotion-neutral stimulus, i.e., the picture of a can of coke. This time, however, the picture of a cockroach was quickly flashed before the can of coke. More precisely, the picture of the cockroach was displayed for a number of milliseconds sufficient for Mary's visual system detecting it, but not for Mary being aware of it. The presentation of such a subliminal stimulus had the following two effects. First, it triggered in Mary the pattern of bodily changes related to disgust. Second, even if the picture of the can of coke didn't generate any significant bodily change in Mary, she reported to feel disgust for the can of coke. In order to control whether Mary was really disgusted by the can of coke, and did not merely report that, a further behavioral task was performed: Mary was repeatedly presented with the mug whose picture was displayed in the second trial and with the can of coke whose picture was presented in the third trial, and she was told that she was free to act on each of these objects, if she wanted to do so. Importantly, while Mary reached for the cup and manipulated it many times, she consistently showed an avoidance behavior towards the can of coke.

These two stories get Prinz into trouble. Without any doubt, Eric was angry at Alex, and not at Pete the Runt. Indeed, not only Eric assaulted Alex, but he was also in the dark about Pete the Runt being behind the counter. As far as Eric knew, Alex was the only individual who could have uttered those obscenities to Louise. However, Prinz maintains that an emotion is a perception of bodily changes, and that it is directed towards the particular object that has caused these bodily changes. Therefore, Prinz has to say that Eric's anger was directed towards Pete the Runt, since it was Pete the Runt's behavior that caused Alex's bodily changes. The same holds for the second story. Mary felt disgust for the can of coke. Indeed, not only did she verbally report such a feeling towards the can of coke, but she also showed a consistent pattern of avoidance behavior, which is best explained by the fact that she found the can of coke disgusting. However, it wasn't the visual perception of the can of coke that triggered Mary's bodily changes. In fact, they were triggered by the picture of the cockroach. Hence, Prinz's theory wrongly predicts that Mary felt disgust for the cockroach.

Abstracting from these particular cases, the problem for Prinz is that, even if emotions are often directed towards the particular object that has caused the relevant bodily changes, there are cases in which what counts as the particular object of one's emotional episode isn't what has caused the relevant bodily changes. Prinz's second proposal cannot account for this latter kind of cases.

3.4.3. Third Proposal: Unification

Prinz's third proposal about the relation between emotions (i.e., bodily perceptions) and representations of particular objects moves from the recognition that, in one's emotional experience, the two are deeply entrenched: "We say 'Jones was frightened of the snake', and ... 'Jones was afraid that the snake would bite.' ... In these cases, we cannot fully separate the emotion from its particular object. Fear-of-snakes and fear-that-the-snake-will-bite seem to comprise unified wholes" (Prinz 2004, 179-180). In order to explain the inseparability of emotions from their particular objects, Prinz develops the following account:

When one is angry about an insult, one's thought about that insult is not merely a cause of one's anger, it is part of one's anger. This can be explained by saying that ... emotions contain both embodied appraisals ... and representations of objects or states of affairs. These two components are bound together in the mind. ... However this binding is achieved, it must obey the right dependency relation. When an ... emotion arises, the embodied appraisal must be caused by the representation of the particular object. (Prinz 2004, 181)

In a nutshell, Prinz is now suggesting that, if a somatic theorist wants to account for emotions being directed towards particular objects, she has to maintain that: (i) emotions are complex states that contain both bodily perceptions and representations of particular objects; and (ii) the two components, i.e., bodily perceptions and representations of particular objects, should obey a relation of causal dependency (the latter is the cause of the former).

This proposal faces the following two problems. The first one is the same problem faced by Prinz's second proposal. Consider the case of Pete the Runt again. Clause (ii) states that the representation of the particular object that is part of one's emotion has to be what has caused one's bodily changes. Since Pete the Runt's behavior was the only element that causally contributed to Eric's bodily changes, Prinz's theory wrongly predicts that Eric's anger has a representation of Pete the Runt among its constituents, and hence that Eric's anger is directed towards Pete the Runt. However, Eric's anger was not directed towards Pete the Runt, but towards Alex.

The second problem is more macroscopic, and concerns clause (i). Recall that the fundamental tenet of a pure somatic theory *à la* Prinz is that emotions are entirely constituted by perceptions of bodily changes. Prinz, however, is now suggesting that, in order to account for emotional directedness towards particular objects, we have to maintain that emotions are complex states that contain both bodily perceptions and representations of particular objects. Thus, even if Prinz's third proposal were a good account of the directedness of emotions towards particular objects, it would not be a pure somaticist account of the phenomenon, for it is based on abandoning the idea that emotions consist of bodily perceptions alone.

To sum up, we discussed three ways in which Prinz analyzed the sense in which emotions are directed towards particular objects, and we showed that they are problematic. In the next section, we are going to give our last argument against a pure somatic theory of emotion.

3.5. *The Phenomenology of Emotions*

What is it like to undergo an emotion? Prinz maintains that the phenomenology of emotion is exhausted by bodily feelings, and uses such phenomenological "evidence" to support his pure somatic theory. Indeed, he argues that the best explanation of the "fact" that there's nothing more to an emotional experience than bodily phenomenology is that emotions are bodily perceptions. However, one's *modus ponens* is another's *modus tollens*. Emotional experiences involve more than bodily feelings, and this indicates that emotions are not just bodily perceptions. So, at least, we are going to argue in this section.

According to some authors, the first respect in which emotional experiences outstrip bodily feelings is that emotional experiences involve *feelings towards*, i.e., feelings directed towards "the world beyond the

body” (Goldie, 2002). Imagine an explorer that sees that a lion is approaching her. In such a situation, the explorer experiences different bodily feelings—the feeling of her heart racing fast, the sensation of her hands sweating, etc. In addition, the explorer feels something towards the lion too. Namely, she *feels afraid of the lion*. This latter feeling is a feeling towards, and it’s not clear how it can be reduced to bodily feelings.

Emotional experiences are also said to involve *cognitive phenomenology*, i.e., the phenomenal character associated to cognitive episodes (Horgan and Tienson, 2002; Pitt, 2004). Consider this passage from Kriegel (forthcoming):

The phenomenology of frustration and the phenomenology of indignation certainly ... involve bodily sensations, but they also involve more than that, including ... an experienced cognitive appreciation of a failure (for frustration) or of an injustice (for indignation). Indeed, it is arguable that the somatic phenomenology of frustration and indignation is in fact indistinguishable, and the only phenomenal element that separates feeling disappointed from feeling indignant is this cognitive-phenomenal element.

We can rephrase Kriegel’s point as follows: at least certain emotions—namely higher cognitive emotions such as indignation, shame, pride, etc.—involve the phenomenal character associated to the cognitive episode(s) that is (are) specific of each of them. For example, the phenomenology of indignation involves the phenomenal character associated with the belief that there has been an injustice, while the phenomenology of guilt involves the phenomenal character associated with the belief that there has been a harmful transgression.

Does the fact that emotional episodes involve feelings towards and cognitive phenomenology tell against a pure somatic theory? Apparently, the only way in which Prinz might try to reconcile his pure somatic theory with the existence of such a non-bodily phenomenology is this: even if emotions are nothing but bodily perceptions, they involve non-bodily phenomenology because they are accompanied by other mental states. For example, when the explorer faces the lion, she has many mental states, among which fear (i.e., a bodily perception) and the perceptual representation of the lion. The fact that she is in a state of fear explains why she experiences certain bodily feelings, while the fact that she has a perceptual representation of the lion explains why she has a feeling towards the lion. Analogously, even though guilt is nothing but a bodily perception, it is caused by the belief that one has committed a harmful transgression, and such a belief usually accompanies guilt. It’s the presence of this concomitant belief that explains why guilt involves a certain cognitive phenomenology.

This proposal, however, does not account for the particular way in which bodily feelings, feelings towards and cognitive phenomenology hang together in emotional experiences. In order to introduce the notions of bodily feeling and cognitive phenomenology, we described emotional

experiences as fragmented experiences constituted by different kinds of simultaneous but disjointed phenomenal states. However, this was just an expository convenience. Emotional experiences are in fact phenomenally unified. More precisely, in emotional experiences, bodily feelings, feelings towards and cognitive phenomenology are *phenomenally integrated*. For example, when the explorer faces the lion, it's not the case that she has two co-occurrent but disjointed experiences, i.e., a set of bodily feelings, on the one hand, and a feeling towards the lion, on the other hand. Rather, the explorer undergoes a complex, integrated conscious experience.

The problem for Prinz is that it's hard to see how this kind of phenomenal integration can result from the mere co-occurrence of different mental states. Consider guilt again. If guilt (i.e., a bodily perception) and the belief that there has been a harmful transgression are two distinct, co-occurring mental states, then we should expect that the typical phenomenology of guilt is just a case of co-consciousness. That is, Prinz's proposal predicts that the phenomenology of guilt is constituted by the mere juxtaposition of bodily feelings and cognitive phenomenology. But this is a mischaracterization of the phenomenology of guilt. In fact, to experience guilt is to experience an integrated, complex phenomenal experience in which bodily feelings and cognitive phenomenology are merged together. Hence, a pure somatic theory cannot do justice of the unity of emotional phenomenology.

3.6. *Taking Stock*

In this section, we've argued that Prinz is right in favoring a somatic theory of emotion over a non-somatic one. Indeed, there is strong evidence in support of the claim that bodily changes causally contribute to emotions, and not the other way around. On the other hand, we've also showed that Prinz's pure somatic theory faces the following difficulties:

- (1) It gets the neural correlates of emotion wrong.
- (2) It does not distinguish between emotions and interoceptive states that are not emotions.
- (3) It cannot account for the fact that emotions are directed towards particular objects.
- (4) It cannot account for emotion phenomenology.

In the next section, we are going to propose an impure somatic theory of emotion that, while accounting for the fact that bodily changes causally contribute to emotions, can also solve the problems the beset a pure somatic theory.

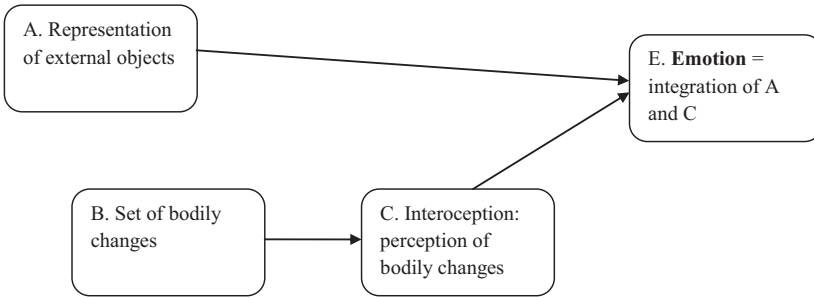


Figure 2. According to our impure somatic theory of emotion, the cognitive process that gives rise to an emotion takes two different types of states as input: (A) representations of external objects/events/states of affairs and (C) interoceptive states, i.e., perceptions of bodily changes. An emotion, i.e., (E), is the result of the integration between (A) and (C).

4. An Impure Somatic Theory of Emotion

This last section is organized as follows. First, we'll sketch the outline of our impure somatic theory against the background of Prinz's pure somatic one (§4.1). Second, we shall show how our theory solves the problems faced by Prinz's theory, and, in doing so, we'll illustrate the details of our proposal (§4.2). Finally, we shall discuss how our theory accounts for the evidence considered by Prinz (§4.3).

4.1. Our Theory in a Nutshell

Simplifying a little, Prinz's account of the nature of emotions is as follows: an emotion E is the perception of the set of bodily changes that are caused by the representations contained in the E calibration file—where the fact that such bodily changes are caused by the representations contained in the E calibration file is due either to evolution or to learning. For example, evolution set in place that perceptual representations of coiled objects trigger a certain set of bodily changes, the perception of which constitutes fear. Analogously, learning set in place that representations of guns trigger the same set of bodily changes, and also in this case the interoception of such bodily changes constitutes fear.

On the other hand, our theory of emotion can be depicted as follows (Figure 2).¹⁹ According to our proposal, emotions, i.e., (E), are constituted by the *integration* of two different types of state: (C) interoceptive states and (A) representations of external objects (events, or states of affairs). In order to have a better understanding of our theory, let's examine the

¹⁹ As we will see in section 4.2.3, this is a simplified schema, which works for basic emotions, but has to be supplemented in order to capture the nature of higher cognitive emotions.

following two questions. First, what's the relation between (B) bodily changes and (A) representations of external objects? Second, what's the nature of the integration processes that combine (C) interoceptive states and (A) representations of external objects together?

As to the first question, there are cases in which the relevant bodily changes are *caused by* the representation of external objects that will be subsequently integrated with the perception of such bodily changes to give rise to one's emotional state. Consider this simple case. An individual has the perceptual representation of a snake. Since evolution set in place that this type of representation causes a certain set of bodily changes, the representation of the snake triggers such bodily changes. The interoceptive system registers these bodily changes. When this interoceptive state integrates with the perceptual representation of the snake, fear occurs. On the other hand, there are cases in which the relevant bodily changes are *causally independent of* the representation of external objects that combines with the interoception of them to generate an emotional state. For example, this is what happens in the case of Pete the Runt. As we will argue in the next section, Eric's emotional state is constituted by the integration of a perception of bodily changes with the visual representation of Alex. However, what triggered Eric's bodily changes was not the visual representation of Alex, but the auditory perception of the obscenities uttered by Pete the Runt.

In any case, according to our theory, emotions are constituted by the integration of interoceptive states with representations of external objects. Thus, we depart from Prinz's pure somatic theory in that we deny that emotions are constituted by interoceptive states alone. This is why our theory counts as an *impure somatic theory of emotion*: because it maintains that bodily changes causally contribute to emotions, but denies that the perception of such bodily changes is all that there is to the constitution of emotions.

Let's turn to the second question: what is the nature of the integration processes that put together interoceptive states and representations of external objects? This is a tough question, and at the present stage we do not have a full-fledged answer to it. However, it's possible to speculate that two kinds of integration processes are involved in emotion generation. First, emotion can involve *perceptual integration*. For example, when a perceptual representation of a snake integrates with certain bodily perceptions, the resulting emotional state, e.g., fear, might be considered as a *multimodal perceptual state*. Second, emotion generation can depend on *cognitive integration*. Take this case. John thinks about tomorrow's exam; this cognitive state triggers a set of bodily changes in John; when the interoception of such bodily changes integrates with the cognitive state that triggered them, John undergoes fear. Importantly, an emotional episode can involve both

types of informational integration. For example, an episode of fear can result from the integration of the following states: the visual perception of an examination paper (perceptual representation), the belief that you are not prepared for the exam (cognitive representation), and the interoception of a set of bodily changes.

Now that you know the general outline of our impure somatic theory, we can move to illustrating its finer details and its explanatory power.

4.2. Problems Solved

In section 3, we argued that Prinz's pure somatic theory faces different problems, namely problems (1)-(4), briefly summarized in section 3.6. In this section, we show that our theory has the explanatory resources to solve these problems. In so doing, we'll also clarify the details of our proposal.

4.2.1. The Neural Correlates of Emotions

Let's start with problem (1). If, as Prinz claims, emotions are entirely constituted by interoceptive states, then the neural correlates of emotions should coincide with the neural correlates of interoception. This is not the case, however. In section 3.2, for example, we showed that, even though both disgust and interoception correlate with insular activity, disgust is "located" in the anterior insula, while the insular correlate of interoception is the posterior insula.

On the other hand, if emotions are constituted by the integration of interoceptive states with representations of external objects, then the neural correlates of emotions should be the loci of these integration processes. Disgust perfectly fits with this hypothesis. Indeed, the neural correlate of disgust (i.e., the anterior insula) is the brain area where the interoceptive representations coming from the posterior insula are integrated with information coming from many brain regions, including those responsible for elaborating sensory information about the external world (Mufson and Mesulam, 1982; Craig, 2002; Harrison et al., 2010).

4.2.2. Representing Particular Objects

Let's now turn to problem (3).²⁰ Emotional episodes are directed towards particular objects, and this is usually accounted for in terms of emotions representing them. However, this analysis of emotional directedness is not available to a pure somatic theorist. For this reason, Prinz proposes that emotions are directed towards particular objects because they are *linked* to other mental states that represent those particular objects. As we have seen

²⁰ We tackle problem (2) in section 4.2.3.

in section 3.4, however, such a proposal faces many difficulties. On the other hand, our impure somatic theory has the resources to account for emotional directedness towards particular objects in terms of emotions representing them. In this section, we cash out such an account.

According to our theory, emotions are constituted by the integration of (C) interoceptive states with (A) representations of external objects. We maintain that the role played by (A) is exactly that of providing an emotion with its particular object. Take this case. I see a snake in my vicinities; this perceptual representation triggers a set of bodily changes; I interocept these bodily changes. When this interoceptive state gets integrated with the perceptual representation of the snake, I am in a state of fear. Why is my fear directed towards that snake? Why does my fear represent that snake? (In our theory, there's no difference between these two questions, since we account for directedness in terms of representation). This is the answer: my fear is directed towards/represents that snake because it has the perceptual representation of that snake among its constituents.

Importantly, our theory also accounts for what Prinz describes as the unity between emotions and their particular objects: "We say 'Jones was frightened of the snake.' ... In these cases, we cannot fully separate the emotion from its particular object. Fear-of-snakes ... seem to comprise unified wholes" (Prinz, 2004, 179-80). We can straightforwardly make sense of this phenomenon: emotions and their particular objects cannot be separated, because emotions are constituted by the integration of interoceptive states with representations of such particular objects. In other words, an emotion and the particular object it is directed towards form a unified whole because the representation of such a particular object is part of the emotion itself.

Finally, our theory can also deal with cases in which an emotion is not directed towards what has caused the relevant bodily changes. For example, this is how we account for the fact that Eric's anger is directed towards Alex, and not towards Pete the Runt. According to our theory, Eric's anger is constituted by the integration of two different states. The first one is the interoception of the bodily changes triggered by Pete the Runt's words. The second component is a representation of external objects. Since Pete the Runt could not be perceptually detected by Eric, the only *salient* representation of external objects available to Eric was the visual perception of Alex. For this reason, Eric's cognitive system selected the visual perception of Alex as an appropriate input for the integration process that resulted in Eric's anger. This is why Eric's anger was directed towards Alex.

Of course, Eric was *wrong* in being angry at Alex. In fact, he should have been angry at Pete the Runt. Alex was entirely innocent. On the other

hand, Eric was *justified* in being angry at Alex, since he was in an epistemic position in which the available evidence strongly supported the conclusion that it was Alex who uttered such obscenities. This divorce between being wrong and being justified offers us the opportunity to make two other points. First, in the good cases, our cognitive system is able to keep track of what has caused the relevant bodily changes, and hence to select for the process of integration the right representation of the external world, i.e., the representation that has caused the relevant bodily changes. However, there are tricky situations in which our cognitive system gets fooled and selects for the process of integration a representation of the external world that has nothing to do with the relevant bodily changes. When this happens, a particular kind of emotional error occurs, namely one's emotion is directed towards the *wrong* object.

The second point is that a subject's emotion can be directed towards the wrong object, even if the subject is *fully rational* and her cognitive system *works properly*. Given Eric's epistemic position, he was justified in being angry at Alex. Moreover, his cognitive system worked fine. Indeed, it did not select a random representation of the external world for the process of integration. Rather, it selected the most salient representation available, namely the visual representation of Alex. So, here is a recipe to fool a subject's emotional system: (i) trigger in the subject a certain representation of the external world, which in turn causes the subject to undergo some emotion-relevant bodily changes; (ii) confound the subject's cognitive system by making it "believe" that another event, which has nothing to do with the undergoing bodily changes, is in fact causally responsible for these bodily changes. As a result, the subject's cognitive system will select the representation of the apparent cause as input for the integration process. Therefore, the resulting subject's emotion will be wrongly directed towards the apparent cause. This is exactly what happens in the story about the disgusting can of coke that we have told in section 3.4.2.

4.2.3. Representing Core Relational Themes

In the previous section, we have considered how our theory accounts for emotions representing particular objects. This, however, is only one aspect of the intentionality of emotions. In fact, we agree with Prinz and Lazarus that emotions represent core relational themes. In this section, we are going to illustrate how our theory accounts for this. To have a general sense of our proposal, let's start by briefly comparing it with Prinz's and Lazarus's.

According to Lazarus (1991), the ability of emotions to represent core relational themes depends on concept possession, for he maintains that

emotions represent core relational themes through *appraisal judgments*. For example, fear represents danger because it involves a judgment (e.g., *I am facing a danger*) whose content has the concept DANGER among its constituents. On the other hand, Prinz maintains that Lazarus's account over-intellectualizes emotions, and, drawing on Dretske's psychosemantics, he proposes that emotions represent core relational themes simply in virtue of being reliably caused by, and having been set in place to carry information about, them. What about us? We think that there is a grain of truth in both positions. In what follows, indeed, we draw on Lazarus's ideas to account for how higher cognitive emotions represent core relational themes, while we resort to Prinz's proposal to explain how basic emotions do so.

Basic emotions. There's remarkable evidence in support of the claim that basic emotions are not the privilege of human beings, but can be experienced by many non-human animals with a rudimentary conceptual apparatus, or even possessing no concept at all (LeDoux, 1996; Panksepp, 1998; 2005). This suggests that, *pace* Lazarus, basic emotions do not represent core relational themes through appraisal judgments.²¹ Thus, how do they do it? This time, our answer is the same as Prinz's. Recall that there is evidence that each type of basic emotion involves a specific body state pattern (Levenson et al., 1990). Since we defend a somatic theory, we interpret this evidence as follows: each type of basic emotion is reliably caused by a specific body state pattern. Such a specific body state pattern is in turn reliably caused by a specific core relational theme. Therefore, each type of basic emotion is reliably caused by a specific core relational theme, and hence carries information about it. Moreover, it's a plausible hypothesis that each type of basic emotion has been set in place to carry such information. Thus, each type of basic emotion represents a specific core relational theme because it is reliably caused by it and has been set in place to carry information about it.

²¹ This, of course, does not mean that the possession of a concept like DANGER plays no important role in a subject's affective and cognitive economy. For example, in virtue of having such a concept, a subject can form the *belief* that she is facing a danger, and she can use such a belief to *deliberate* about what to do. The concept DANGER is also important for *emotion regulation*. Indeed, if a subject is experiencing fear and *believes* that she is afraid of something that is not in fact dangerous, she can try to exert some cognitive control over her affective state. However, the fact that a creature endowed with the concept DANGER has more cognitive resources than a creature who lacks such a concept does not ground the conclusion that the basic emotions of these two kinds of creature represent core relational themes in a radically different way. In fact, it is more plausible to maintain that basic emotions are mental commodities that allow a creature to represent a core relational theme without the need of representing it *as such*, independently of whether such a creature possesses the relevant concepts to do so.

Importantly, we can subscribe to Prinz's account of how basic emotions represent core relational themes without facing problem (2), i.e., the inability to distinguish emotions from non-emotional states. Recall that Prinz endorses both the following thesis:

Distinctiveness Thesis (DT): Emotions are interoceptive states, and what distinguishes emotions from non-emotional interoceptive states is that emotions are those interoceptive states that represent core relational themes.

Psychosemantics thesis (PT): A mental state represents a core relational theme if it is reliably caused by it and has been set in place to carry information about it.

As we know, the conjunction of (DT) and (PT) has an undesirable consequence. Take the feeling of ear pressure. This is an interoceptive state that is reliably caused by a core relational theme, namely finding oneself in an environment with unhealthy degrees of barometric pressure and oxygen levels, and that has been set up by evolution to carry information about such a core relational theme. Therefore, (PT) predicts that the feeling of ear pressure is an interoceptive state that represents a core relation theme. However, if one also subscribes to (DT), then one has to conclude that ear pressure is an emotion, which is clearly absurd. On the contrary, our theory does not face this problem. In fact, even if we endorse that (PT) is the right explanation of how basic emotions represent core relational themes, we do not subscribe to (DT), for we deny that emotions are identical to interoceptive states. Therefore, even if both fear and the feeling of ear pressure represent core relational themes, we have a principled way to draw a distinction between them.

Higher cognitive emotions. Let's now consider how higher cognitive emotions represent core relational themes. This time, our proposal is by far more tentative, since we do not have a full-fledged account of the nature of higher cognitive emotions. We provisionally endorse the idea that basic emotions and higher cognitive emotions differ in at least two respects. First, any type of higher cognitive emotion has one or more *type-specific propositional attitudes* among its constituents. For example, guilt is partly constituted by the belief that one has committed a harmful transgression (or a relevantly similar propositional attitude), while jealousy is partly constituted by the belief that someone is a threat for one's affective relations (or a relevantly similar propositional attitude). To put it in mechanistic terms, the idea is that the integration process that gives rise to a higher cognitive emotion always takes a certain type-specific propositional attitude as input. Take this case. Mary sees Liza flirting with her husband

John. On this basis, she forms the belief that Liza is a threat to her marriage. These visual and cognitive states trigger a set of bodily changes in her. When (A) the visual perception of Liza (D) the belief that Liza is a threat to her marriage and (C) the interoception of the bodily changes triggered by (A) and (D) get integrated, Mary undergoes jealousy (The process giving rise to higher cognitive emotions is depicted in Figure 3).

In the next section, when we will discuss emotion phenomenology, we will give an argument in support of the claim that higher cognitive emotions have such type-specific propositional attitudes among their constituents. Now, let's consider the second difference between basic emotions and higher cognitive emotions. We have claimed that each type of basic emotion is reliably caused by a specific body state pattern, and we drew on it to account for how basic emotions represent core relational themes. On the other hand, it is far from clear that there exist specific body state patterns for higher cognitive emotions. Indeed, empirical evidence does not rule out that different types of higher cognitive emotions might be associated with the same bodily changes.²² If so, higher cognitive emotions cannot represent core relational themes in the same way as basic emotions do. An alternative mechanism is needed.

Recall that higher cognitive emotions have type-specific propositional attitudes among their constituents. For example, jealousy is partly constituted by the belief that someone is a threat for one's affective relations, while guilt is partly constituted by the belief that one has committed a harmful transgression. Importantly, the contents of these type-specific propositional attitudes are core relational themes. Indeed, the core relational theme of jealousy is that someone is a threat for one's affective relations, while the core relational theme of guilt is that one has committed a harmful transgression (Lazarus 1991). Therefore, what we have called 'type-specific propositional attitudes' are nothing but Lazarus's (1991) *appraisal judgments*. Accordingly, we propose an account of the intentionality of higher cognitive emotions along Lazarus's line: higher cognitive emotions represent core relational themes in virtue of being partly constituted by appraisal judgments, i.e., propositional attitudes whose contents are core relational themes.

²² However, Clark (2010) suggests not to jump too quickly to a conclusion. In fact, he argues that the lack of evidence in support of the claim that higher cognitive emotions have type-specific body state patterns might just be due to the fact that their physiological correlates haven't been studied extensively. Moreover, he maintains that shame is associated with a specific body change, namely inflammatory immune system responses. We will not discuss here whether such immune responses are shame-specific. In any case, we agree with Clark that the issue of whether higher cognitive emotions have type-specific body state patterns requires more investigation.

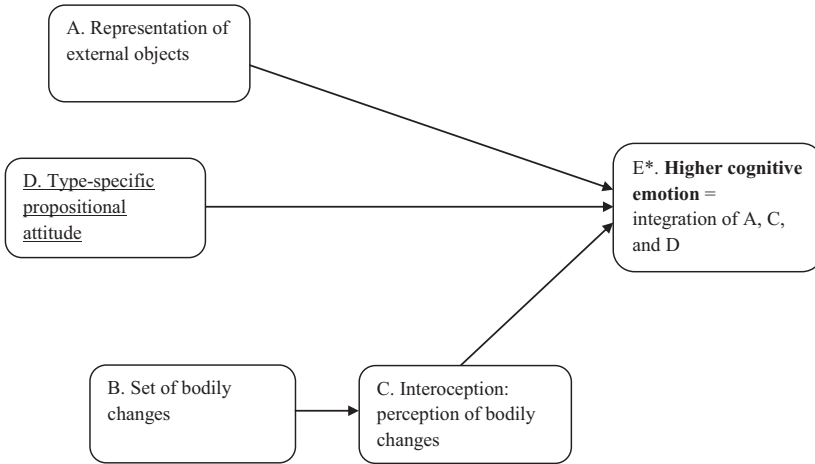


Figure 3. Higher cognitive emotions, (E*), are constituted by the integration of three types of mental states: (A) representations of external objects, (C) interoceptive states, and (D) type-specific propositional attitudes.

Is emotion a natural kind? At this point, one might legitimately ask what our account of the differences between basic emotions and higher cognitive emotions implies relative to the issue as to whether emotion is a natural kind. Does our theory force us to endorse the disunity thesis, or is it compatible with the claim that basic emotions and higher cognitive emotions are members of the same natural kind? One problem in answering this question is that there is no agreement on the criteria for classifying natural kinds (see Bird and Tobin, 2012). As we have said in section 2.3, Prinz considers two accounts of natural kinds: (i) *Essentialism*, according to which members of a natural kind share a common essence (Kripke, 1980; Putnam, 1975); (ii) *Homeostatic Property Cluster Theory*, which maintains that a natural kind is a class of individuals that tend to share a cluster of properties because of the operations of a homeostatic mechanism (Boyd, 1991; 1999). Prinz claims that his theory of emotion makes emotion a natural kind in both senses. What about ours?

Let's begin by considering whether basic emotions and higher cognitive emotions share a common essence. In Prinz's theory they do, because both basic emotions and higher cognitive emotions are entirely constituted by bodily perceptions only. On the other hand, we have proposed that, unlike basic emotions, higher cognitive emotions have appraisal judgments among their constituents. If we are right, emotion probably does not qualify as a natural kind in the essentialist sense.

Let's now move on to the question as to whether by our theory emotion constitutes a natural kind in the weaker, Boydian sense. We take it

that, if emotion is a natural kind, it is a *psychological kind*. Hence, the question amounts to asking whether emotion possesses the two following characteristics: (1) its members tend to share a large set of properties; (2) they tend to share these properties because of the operations of a common *cognitive mechanism*. Now, in our account, basic emotions and higher cognitive emotions do share a large set of properties. For example, they both involve bodily changes and bodily perceptions; they both represent particular objects and core relational themes; both are associated with bodily feelings and feelings towards (see the next section); both are motivational states, etc. Therefore, emotion satisfies requirement (1). Moreover, we have stressed that both basic emotions and higher cognitive emotions are the result of the operations of a mechanism of cognitive integration. This, indeed, was the central aspect of our proposal. On the basis of this, one might be tempted to conclude that emotion also satisfies requirement (2), and thus that emotion is a natural kind in the Boydian sense. This conclusion, however, would be premature. In fact, while basic emotions are constituted by the integration of bodily perceptions with representations of external objects, higher cognitive emotions require the integration of a further element (i.e., appraisal judgments), and, at the present stage, there is not enough empirical evidence for the claim that the cognitive mechanism that performs these integrations is the same in both cases. Therefore, on the basis of our theory and of the currently available empirical evidence, we can only make the following conditional statement about emotion being a natural kind (in the Boydian sense): if it turns out that the same mechanism of cognitive integration is in place both in the case of basic emotions and in the case of higher cognitive emotions, then emotion is a Boydian natural kind.²³

²³ Interestingly, similar considerations also apply to our treatment of basic emotions. As we have said in section 4.1, a token basic emotion can be generated in two different ways: by the integration of a bodily perception with the perceptual representation of an external object/event (e.g., the visual perception of a snake), or by the integration of a bodily perception with the cognitive representation of an external object/event (e.g., the belief that tomorrow there will be an exam). This suggests that basic emotion does not qualify as a natural kind in the essentialist sense. In fact, members of the basic emotion category do not share a common essence, but splinter into multimodal perceptual states and cognitive states. The former are constituted by the integration of bodily perceptions with perceptual representations of the external world; the latter are constituted by the integration of bodily perceptions with cognitive representations of the external world (e.g., beliefs). Is basic emotion a natural kind at least in the Boydian sense? Again, to settle this issue, further empirical evidence is needed. If it turns out that the integration mechanism that puts together bodily perceptions and cognitive representations of the external world is the same mechanism that is causally responsible for integrating bodily perceptions with perceptual representations of the external world, then basic emotion is a Boydian kind. If, on the other hand, these are two distinct and independent mechanisms, basic emotion splinters into two incongruous categories.

4.2.4. Emotion Phenomenology

Finally, we discuss the last problem that afflicts Prinz's theory, i.e., problem (4): the inability to account for the fact that emotion phenomenology is constituted by the integration of different types of feeling. Let's begin by illustrating how our theory explains the integration of bodily feelings with feeling towards. Consider again the case of the explorer that sees that a lion is approaching her, gets scared, and undergoes an integrated conscious experience in which certain bodily feelings and the feeling of being afraid of the lion are merged together. According to our theory, the explorer's fear is constituted by the integration of two states, namely (i) the visual representation of the lion and (ii) a certain set of bodily perceptions. Presumably, (i) and (ii) are conscious states, that is, the explorer undergoes a certain *visual experience* of the lion and a certain set of *bodily feelings*. Therefore, when (i) and (ii) get integrated, we do not simply have a case of informational integration, but also a case of *phenomenal integration*: even though the visual experience of the lion and the set of bodily feelings belong to two different sense modalities, i.e., vision and interoception, they get integrated in a single phenomenal experience, and this accounts for the peculiar unity of the explorer's experience.

Similar considerations apply to cognitive phenomenology. We have proposed that higher cognitive emotions essentially involve appraisal judgments among their *constituents*. This proposal contrasts with Prinz's idea that higher cognitive emotions are entirely constituted by bodily perceptions and involve propositional attitudes only among their *causes*. The phenomenology of higher cognitive emotions shows that our proposal has to be favored over Prinz's. Consider guilt. If guilt is just a bodily perception which is caused and accompanied by the belief that one has committed a harmful transgression, it should be the case that the phenomenology of guilt is constituted by the mere juxtaposition of bodily feelings and the cognitive phenomenology associated with the belief that one has committed a harmful transgression. However, this is a mischaracterization of the phenomenology of guilt. In fact, the experience of guilt is constituted by the integration of bodily feelings with the cognitive phenomenology associated with the belief that one has committed a harmful transgression. Our theory can account for this. Guilt has such a phenomenal character because it is constituted by the integration of bodily perceptions with the belief (appraisal judgment) that one has committed a harmful transgression.

4.3. Explaining the Evidence Considered by Prinz

In the previous section, we have showed how our impure somatic theory can solve the problems that beset Prinz's pure somatic theory. In this final

section, we are going to argue that our theory can also account for the evidence that Prinz's pure somatic theory was introduced to explain, namely:

- i. Each type of basic emotion has its own pattern of associated bodily changes (Levenson et al., 1990).
- ii. Neuroimaging studies revealed activation in the neural correlates of interoception, i.e., somatosensory and insular cortex, during emotional episodes (Damasio et al., 2000).
- iii. Changes in facial expressions influence one's emotional reaction towards eliciting emotional stimuli (Strack et al., 1988; Zajonc et al., 1989).
- iv. Neuropsychological evidence shows that damages to insular cortex cause deficits in emotional experience (Hennenlotter et al., 2004).

To begin with, if emotions are constituted by the integration of interoceptive states with representations of external objects, then the neural correlates of interoception have to be recruited during emotion processing (evidence ii), even though they are not the neural correlates of emotions (see section 4.2.1). In addition, our theory predicts that the disruption of interoception gives rise to emotional deficits (evidence iv), since interoceptive states are one the constituents of emotions. Moreover, we have argued that each type of basic emotion represents its core relational themes in virtue of being reliably caused by a specific body state pattern. Therefore, our theory is committed to the existence of specific body state patterns for each type of basic emotion (evidence i).

Finally, let's consider evidence ii, which shows that voluntary and involuntary changes in people's facial expressions influence their emotional reactions towards external stimuli. For example, Strack et al. (1988) found that subjects with a pen between their teeth rated a cartoon as funnier in comparison with subjects who held a pen between their lips. A case like this can be easily accounted for by our theory. We maintain that emotions are constituted by the integration of interoceptive states with representations of external objects. In both conditions, subjects had the same representation of external objects, namely the representation of the cartoon. However, while subjects in the teeth condition were smiling, subjects in the lip condition were grimacing. Therefore, subjects in the teeth condition underwent a different interoceptive state than subjects in the lips condition. This explains why the two groups of subjects had different emotional experiences.

To conclude, an impure somatic theory has the following explanatory advantage over Prinz's pure somatic theory: it can easily explain the

experimental evidence considered by Prinz and, at the same time, it can solve all the problems that afflict Prinz's theory. Therefore, if you are interested in a somatic theory of emotion, we warmly invite you to buy an impure somatic version of it.^{24,25}

²⁴ In order to have a better grasp of our proposal, it can be useful to briefly contrast it with another influential contemporary somatic theory of emotion, i.e., the one developed by Damasio (1994), which, for reasons of space, we have neglected in this article. Damasio's theory is very complex, and we cannot discuss it in detail here. We mention, however, four points on which Damasio's theory and our theory diverge.

The *first* point concerns the nature of emotion, i.e., *what emotions are*. In our account, emotions are mental states constituted by the integration of bodily perceptions with representations of external objects (*plus* cognitive appraisals, in the case of higher cognitive emotions). According to Damasio, on the other hand, "the essence of emotion is a collection of changes in body states" (Damasio, 1994, 139). This also distinguishes Damasio's theory from Prinz's. In fact, while Prinz maintains that emotions are perceptions of bodily changes, Damasio claims that emotions *are bodily changes*, and he explicitly "leaves out of emotion the perception of the changes that constitute the emotional response" (Damasio, 1994, 139). Thus, strictly speaking, for Damasio emotions are not mental states, but bodily states. A *second* difference between our theory and Damasio's concerns the *causal process* that brings about emotions. According to Damasio, this process "begins with the conscious, deliberate considerations you entertain about a person or situation ... , [i.e., it begins with] a cognitive evaluation of ... the event of which you are a part" (Damasio, 1994, 139). In a nutshell, Damasio maintains that emotions are *always* preceded by a process of *cognitive appraisal*. We disagree. In fact, even though we endorse the idea that higher cognitive emotions are caused (among other things) by cognitive appraisals, we deny that basic emotions need to involve a process of cognitive appraisal (see section 4.2.3 of this paper).

One could maintain, however, that there is at least one important point on which we agree with Damasio, namely, the nature of *emotion phenomenology*. Damasio writes: "The [emotional] process does not stop with the bodily changes that define an emotion. The ... next step is the *feeling of the emotion* in connection to the [external] object that excited it, the realization of the nexus between object and emotional body state" (Damasio, 1994, 132). Isn't this what we proposed in section 4.2.4, when we argued that feelings towards are due to the *phenomenal integration* of bodily feelings with conscious representations of external objects? It isn't—and this is the *third* difference with Damasio's theory. Damasio claims: "the essence of feeling an emotion is the experience of ... [bodily] changes in *juxtaposition* to the mental images [of external objects] that initiated [such bodily changes]" (Damasio, 1994, 145). Thus, while we maintain that emotion phenomenology is such that bodily feelings are *merged with* conscious representations of external objects, Damasio maintains that there is no blending, but simply co-occurrence.

Finally, we disagree with Damasio in the way of conceiving the cognitive/neural *process underlying emotional experience*: we maintain that this process is such that bodily perceptions and representations of external objects get cognitively/neurally integrated; on the other hand, Damasio thinks that "the image of the body proper [and the image of the external object] remain separate, neurally" (Damasio, 1994, 146).

²⁵ Earlier versions of this article were presented at the Institut Jean Nicod, at the Ruhr-Universität Bochum, at the University of Turku, and at the 2012 Milan-Nottingham-York workshop, where we received many thought-provoking questions and criticisms. We warmly thank the following people, who provided numerous insightful comments: Kenneth Aizawa, Matteo Baccharini, Claudio Bergamini, Davide Bordini, Helen Bradley, Chiara Brozzo, Clotilde Calabi, Lena Kästner, Felipe Nogueira de Carvalho, Raphael van Riel, Pietro Snider, Anna Welpinghus, and an anonymous reviewer. Special thanks go to Kevin

References

- Bird, Alexander and E. Tobin “Natural Kinds.” *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2012/entries/natural-kinds/>>.
- Boyd, Richard. 1991. “Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds.” *Philosophical Studies* 61: 127–148.
- . 1999. “Homeostasis, Species and Higher Taxa.” In R. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press, 141–185.
- Cannon, Walter B. 1927. “The James-Lange Theory of Emotion: A Critical Examination and an Alternative Theory.” *American Journal of Psychology* 39: 106–124.
- Clark, Jason A. 2010. “Relations of Homology between Higher Cognitive Emotions and Basic Emotions.” *Biology and Philosophy* 25(1): 75–94.
- Craig, A. D. 2002. “How Do You Feel? Interoception: the Sense of the Physiological Condition of the Body.” *Nature Reviews Neuroscience* 3 (8): 655–666.
- . 2003. “Interoception: The Sense of the Physiological Condition of the Body.” *Current Opinion in Neurobiology* 13(4): 500–505.
- . 2009. “How Do You Feel—Now? The Anterior Insula and Human Awareness.” *Nature Reviews Neuroscience* 10(1): 59–70.
- Damasio, Antonio. (1994). *Descartes’ Error: Emotion, Reason and the Human Brain*. New York: Putnam.
- , T. Grabowski, A. Bechara, H. Damasio, L.L. Ponto, J. Parvizi and R.D. Hichwa. 2000. “Subcortical and Cortical Brain Activity during the Feeling of Self-Generated Emotions.” *Nature Neuroscience* 3(10): 1049–1056.
- , H. Damasio and D. Tranel. Forthcoming. “Persistence of Feelings and Sentience after Bilateral Damage of the Insula.” *Cerebral Cortex*.
- David, Wayne. 1988. “A Causal Theory of Experiential Fear.” *Canadian Journal of Philosophy* 18(3): 459–483.
- Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge: MIT Press.
- . 1986. “Misrepresentation.” In R. Bogdan (ed.), *Belief: Form, Content and Function*. Oxford: Oxford University Press.

Reuter and Tomoo Ueda, with whom we had invaluable hours-long (sometimes day-long) discussions on the ideas developed in this article. We owe our greatest debt of gratitude to Sandro Zucchi, who takes a second to realize what you didn’t realize in a month and who, when you have already thought it twice, forces you to think it thrice. Finally, LB wishes to thank the Von Humboldt Foundation for generously supporting his research.

- Ellsworth, Phoebe C. 1994. "William James and Emotion: Is a Century of Fame Worth a Century of Misunderstanding?" *Psychological Review* 101(2): 222–229.
- Ekman, Paul. 1972. "Universals and Cultural Differences in Facial Expressions of Emotions." In James Cole (ed.), *Nebraska Symposium on Motivation, 1971, Vol. 19*. Lincoln: University of Nebraska Press, 207–283.
- . 1999. "Basic Emotions." In Dalglish and Power (eds.), *The Handbook of Cognition and Emotion*. New York: Wiley, 45–60.
- Gianaros, Peter J., K.S. Quigley, J.T. Mordkoff and R.M. Stern. 2001. "Gastric Myoelectrical and Autonomic Cardiac Reactivity to Laboratory Stressors." *Psychophysiology* 38(4): 642–652.
- Goldie, Peter. 2002. "Emotions, Feelings and Intentionality." *Phenomenology and the Cognitive Sciences* 1(3): 235–254.
- Goldman, Alvin and F. de Vignemont. 2009. "Is Social Cognition Embodied?" *Trends in Cognitive Sciences* 13(4): 154–159.
- Griffiths, Paul E. 1997. *What Emotions Really Are*. Chicago: University of Chicago Press.
- Harrison, Neil A., M.A. Gray, P.J. Gianaros and H.D. Critchley. 2010. "The Embodiment of Emotional Feelings in the Brain." *The Journal of Neuroscience* 30(38): 12878–12884.
- Hennenlotter, Andreas, U. Schroeder, P. Erhard, B. Haslinger, R. Stahl, A. Weindl, H. G. von Einsiedel, K. W. Lange and A. O. Ceballos-Baumann. 2004. "Neural Correlates Associated with Impaired Disgust Processing in Pre-symptomatic Huntington's Disease." *Brain* 127 (6): 1446–1453.
- Horgan, Terrence E. and J.L. Tienson. 2002. "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Reading*. New York: Oxford University Press.
- Izard, Carroll E. 2007. "Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm." *Perspectives on Psychological Science* 2(3): 260–280.
- James, William. 1884. "What is an Emotion?" *Mind* 9(34): 188–205.
- . 1890. *The Principles of Psychology*. Mineola: Dover Publications.
- Jokerst, M. D., M. Levine, R.M. Stern and K.L. Koch. 1997. "Modified Sham Feeding with Pleasant and Disgusting Foods: Cephalic-Vagal Influences in Gastric Myoelectric Activity." *Gastroenterology* 112: A755.
- Kenny, Anthony. 1963. *Action, Emotion and Will*. London: Routledge.
- Khalsa, Sahib S., D. Rudrauf, J.S. Feinstein and D. Tranel. 2009. "The Pathways of Interoceptive Awareness." *Nature Neuroscience* 12(12): 1494–1496.

- Kriegel, Uriah. Forthcoming. "Towards a New Feeling Theory of Emotion." *European Journal of Philosophy*.
- Kripke, Saul A. 1980. *Naming and Necessity*. Malden: Blackwell Publishers.
- Krolak-Salmon, Pierre, M.-A. Hénaff, J. Isnard, C. Tallon-Baudry, M. Guénot, A. Vighetto, O. Bertrand and F. Mauguière. 2003. "An Attention Modulated Response to Disgust in Human Ventral Anterior Insula." *Annals of Neurology* 55(4): 446–453.
- Lange, Carl. 1885. "The Emotions." In E. Dunlap (ed.), *The Emotions*. Baltimore, MD: Williams & Wilkins.
- Lazarus, Richard S. 1991. *Emotion and Adaptation*. New York: Oxford University Press.
- LeDoux, Joseph E. 1996. *The Emotional Brain*. New York: Simon and Schuster.
- Levenson, Robert W., P. Ekman and W.V. Friesen. 1990. "Voluntary Facial Action Generates Emotion-Specific Autonomic Nervous System Activity." *Psychophysiology* 27(4): 363–384.
- Mufson, Elliot J. and M.M. Mesulam. 1982. "Insula of the Old World Monkey. II: Afferent Cortical Input and Comments on the Claustrum." *The Journal of Comparative Neurology* 212(1): 23–37.
- Ortony, Andrew and T.J. Turner. 1990. "What's Basic about Basic Emotions?" *Psychological Review* 97(3): 315–331.
- Panksepp, Jaak. 1998. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press.
- . 2005. "Affective Consciousness: Core Emotional Feelings in Animals and Humans." *Consciousness and Cognition* 14(1): 30–80.
- Panksepp, Jaak. 2007. "Emotional Feelings Originate Below the Neocortex: Toward a Neurobiology of the Soul." *Behavioral and Brain Sciences* 30 (1): 101–103.
- Prinz, Jesse J. 2000. "The Duality of Content". *Philosophical Studies* 100 (1): 1–34.
- . 2004. *Gut Reactions*. Oxford: Oxford University Press.
- Putnam, Hilary. 1975. "The Meaning of 'Meaning'." In *Philosophical Papers: Vol. 2. Mind, Language and Reality*, 215–271. Cambridge, UK: Cambridge University Press.
- Reisenzein, Rainer, W-U. Meyer and A. Schützwohl. 1995. "James and the Physical Basis of Emotion: A comment on Ellsworth." *Psychological Review* 102(4): 757– 761.
- Rozin, Paul and A. Fallon. 1987. "A Perspective on Disgust." *Psychological Review* 94(1): 23–41.
- , J. Haidt and C.R. McCauley. 2008. "Disgust." In Lewis, Haviland-Jones and Barrett (eds.), *Handbook of Emotions*, 3rd ed. New York: Guilford Press, 757–776.

- Small, Dana M., M.D. Gregory, Y.E. Mak, D. Gitelman, M.M. Mesulam and T. Parrish. 2003. "Dissociations of Neural Representation of Intensity and Affective Valuation in Human Gustation." *Neuron* 39(4): 701–711.
- Solomon, Robert C. 1984. *The Passions: The Myth and Nature of Human Emotions*. New York: Doubleday.
- Stern, Robert M., H.E. Crawford, W.R. Stewart, M.W. Vasey and K.L. Koch. 1989. "Sham Feeding. Cephalic-Vagal Influences on Gastric Myoelectric Activity." *Digestive Diseases and Sciences* 34(4): 521–527.
- Strack, Fritz, L.L. Martin and S. Stepper. 1988. "Inhibiting and Facilitating Conditions of the Human Smile: a Nonobtrusive Test of the Facial Feedback Hypothesis." *Journal of Personality and Social Psychology* 54(5): 768–777.
- Wicker, Bruno, C. Keysers, J. Plailly, J.-P. Royet, V. Gallese and G. Rizzolatti. 2003. "Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust." *Neuron* 40(3): 655–664.
- Zajonc, Robert B., S.R. Murphy and M. Inglehart. 1989. "Feeling and Facial Efference: Implications of the Vascular Theory of Emotion." *Psychological Review* 96(3): 395–416.